CLAIMS

What is claimed is:

1. A method of adaptively managing pages in a cache memory with a variable workload, said method comprising:

maintaining a bit that is set to either a first identifier or a second identifier for every page in the cache memory to indicate whether the bit has short-term utility or long-term utility; and

adaptively varying a proportion of pages marked as said short-term utility and those marked as said long-term utility to increase a cache hit ratio of said cache memory.

2. The method claim 1, further comprising maintaining a temporal locality window such that pages that are re-requested within the window are of short-term utility and pages that are re-requested outside the window are of long-term utility.

3. A method of adaptively managing pages in a cache memory with a variable workload, said method comprising:

defining a cache memory;

organizing the cache memory into a plurality of disjoint lists of pages, wherein said lists comprise first list of top pages, the first list of bottom pages, the second list of top pages, and the second list of bottom pages;

maintaining a bit that is set to either a first identifier or a second identifier for every page

in the cache memory to indicate whether the bit has short-term utility or long-term utility;

ensuring that each member page of the first list of top pages is marked either as short-term utility or as long-term utility, wherein each member page of said first list of bottom pages is marked as short-term utility and each member page of said second list of top pages and said second list of bottom pages is marked as long-term utility; and

maintaining a temporal locality window parameter such that pages that are re-requested within a specified window are of short-term utility and pages that are re-requested outside the window are of long-term utility, wherein the cache memory comprises pages that are members of any of said first list of top pages and said second list of top pages.

4. The method of claim 3, further comprising adaptively varying sizes of said first list of top pages, said second list of top pages, said first list of bottom pages, and said second list of bottom pages, and adaptively varying the temporal locality window parameter in response to a variable workload.

5. The method of claim 3, wherein said first list of top pages and said first list of bottom pages have a variable total length, and said second list of top pages and said second list of bottom pages have a variable total length.

6. The method of claim 3, wherein said step of defining the cache memory comprises defining a cache memory size measured as a number of pages the cache memory can hold.

7. The method of claim 6, further comprising maintaining a total number of pages in the cache memory that are marked as short-term utility to approximately a same size as the cache memory.

8. The method of claim 6, further comprising maintaining a total number of pages in the cache memory that are marked as long-term utility to approximately a same size as the cache memory.

9. The method of claim 3, further comprising labeling each page with bit short-term utility if said page does not exist in either said first list of top pages, said second list of top pages, said first list of bottom pages, or said second list of bottom pages.

10. The method of claim 3, further comprising changing a label of a page from short-term utility to long-term utility only if said page is in said first list of bottom pages.

11. The method of claim 3, wherein said pages in each of said first list of top pages, said first list of bottom pages, said second list of top pages, and said second list of bottom pages appear in an order according to their respective most recent requests, wherein each list comprises a top page and a bottom page, with said top page listing most recent member pages and said bottom page listing least recent member pages, said bottom page in said first list of top pages is always labeled as short-term utility.

12. A program storage device readable by computer, tangibly embodying a program of instructions executable by said computer to perform a method of adaptively managing pages in a cache memory with a variable workload, said method comprising:

defining a cache memory;

organizing the cache memory into a plurality of disjoint lists of pages, wherein said lists comprise first list of top pages, the first list of bottom pages, the second list of top pages, and the second list of bottom pages;

maintaining a bit that is set to either a first identifier or a second identifier for every page in the cache memory to indicate whether the bit has short-term utility or long-term utility;

ensuring that each member page of the first list of top pages is marked either as short-term utility or as long-term utility, wherein each member page of said first list of bottom pages is marked as short-term utility and each member page of said second list of top pages and said second list of bottom pages is marked as long-term utility; and

maintaining a temporal locality window parameter such that pages that are re-requested within a specified window are of short-term utility and pages that are re-requested outside the window are of long-term utility, wherein the cache memory comprises pages that are members of any of said first list of top pages and said second list of top pages.

13. The program storage device of claim 12, further comprising adaptively varying sizes of said first list of top pages, said second list of top pages, said first list of bottom pages, and said second list of bottom pages, and adaptively varying the temporal locality window parameter in response to a variable workload.

14. The program storage device of claim 12, wherein said first list of top pages and said first list of bottom pages have a variable total length, and said second list of top pages and said second list of bottom pages have a variable total length.

15. The program storage device of claim 12, wherein said step of defining the cache memory comprises defining a cache memory size measured as a number of pages the cache memory can hold.

16. The program storage device of claim 15, further comprising maintaining a total number of pages in the cache memory that are marked as 'S" to approximately a same size as the cache memory.

17. The program storage device of claim 15, further comprising maintaining a total number of pages in the cache memory that are marked as long-term utility to approximately a same size as the cache memory.

18. The program storage device of claim 12, further comprising labeling each page with bit short-term utility if said page does not exist in either said first list of top pages, said second list of top pages, said first list of bottom pages, or said second list of bottom pages.

19. The program storage device of claim 12, further comprising changing a label of a page from short-term utility to long-term utility only if said page is in said first list of bottom pages.

20. The program storage device of claim 12, wherein said pages in each of said first list of top pages, said first list of bottom pages, said second list of top pages, and said second list of bottom pages appear in an order according to their respective most recent requests, wherein each list comprises a top page and a bottom page, with said top page listing most recent member pages and said bottom page listing least recent member pages, said bottom page in said first list of top pages is always labeled as short-term utility.

21. A computer system for adaptively managing pages in a cache memory with a variable workload comprising:

a cache memory directory comprising a plurality of disjoint lists of pages, wherein said lists comprise first list of top pages, second list of top pages, first list of bottom pages, and second list of bottom pages;

a bit marker that marks each of said pages to either a first identifier or a second identifier in said cache memory directory to indicate whether the page has short-term utility or long-term utility, wherein each member page of said first list of top pages is marked either as short-term utility or as long-term utility, wherein each member page of said first list of bottom pages is marked as short-term utility, and wherein each member page of said second list of top pages and said second list of bottom pages is marked as long-term utility; and

a temporal locality window parameter, wherein pages that are re-requested within a

specified window parameter are of short-term utility and pages that are re-requested outside the window parameter are of long-term utility, wherein said cache memory directory comprises pages that are members of any of said first list of top pages and said second list of top pages.

22. The computer system of claim 21, further comprising a controller operable for adaptively varying sizes of said first list of top pages, said second list of top pages, said first list of bottom pages, and said second list of bottom pages, and adaptively varying the temporal locality window parameter in response to a variable workload.

23. The computer system of claim 21, wherein said first list of top pages and said first list of bottom pages have a variable total length, and said second list of top pages and said second list of bottom pages have a variable total length.

24. The computer system of claim 21, wherein said cache memory directory comprises a cache memory directory size measured as a number of pages the cache memory directory can hold.

25. The computer system of claim 24, further comprising a controller operable for maintaining a total number of pages in the cache memory directory that are marked as 'S" to approximately a same size as the cache memory directory.

26. The computer system of claim 24, further comprising a controller operable for maintaining a total number of pages in the cache memory directory that are marked as long-term utility to approximately a same size as the cache memory directory.

27. The computer system of claim 21, further comprising a controller operable for labeling each page with bit short-term utility if said page does not exist in either said first list of top pages, said second list of top pages, said first list of bottom pages, or said second list of bottom pages.

28. The computer system of claim 21, further comprising a controller operable for changing a label of a page from short-term utility to long-term utility only if said page is in said first list of bottom pages.

29. The computer system of claim 21, wherein said pages in each of said first list of top pages, said first list of bottom pages, said second list of top pages, and said second list of bottom pages appear in an order according to their respective most recent requests, wherein each list comprises a top page and a bottom page, with said top page listing most recent member pages and said bottom page listing least recent member pages, said bottom page in said first list of top pages is always labeled as short-term utility.

30. A computer system for adaptively managing pages in a cache memory with a variable workload, said method comprising:

means for defining a cache memory;

means for organizing the cache memory into four disjoint lists of pages, wherein said lists comprise first list of top pages, second list of top pages, first list of bottom pages, and second list of bottom pages;

means for maintaining a bit that is set to either a first identifier or a second identifier for every page in the cache memory to indicate whether the bit has short-term utility or long-term utility;

means for ensuring that each member page of the first list of top pages is marked either as short-term utility or as long-term utility, wherein each member page of said first list of bottom pages is marked as short-term utility and each member page of said second list of top pages and said second list of bottom pages is marked as long-term utility; and

means for maintaining a temporal locality window parameter such that pages that are re-requested within a specified window parameter are of short-term utility and pages that are re-requested outside the window parameter are of long-term utility, wherein the cache memory comprises pages that are members of any of said first list of top pages and said second list of top pages.